Quality in Health Care 1994;3:186–192

SF 36 health survey questionnaire: II. Responsiveness to changes in health status in four common clinical conditions

Andrew M Garratt, Danny A Ruta, Mona I Abdalla, Ian T Russell

Abstract

***Objective*—To assess the responsiveness of the SF 36 health survey questionnaire to changes in health status over time for four common clinical conditions.**

***Design*—Postal questionnaires at baseline and after one year's follow up, with two reminders at two week intervals if necessary.**

***Setting*—Clinics and four training general practices in Grampian region in the north east of Scotland.**

***Patients*—More than 1700 patients aged 16 to 86 years with one of four conditions: low back pain, menorrhagia, suspected peptic ulcer, and varicose veins; and a random sample of 900 members of the local general population for comparison.**

***Main measures*—A transition question measuring change in health and the eight scales of the SF 36 health survey questionnaire; standardised response means (mean change in score for a scale divided by the standard deviation of the change in scores) used to quantify the instrument's responsiveness to changes in perceived health status, and comparison of patient scores at baseline and follow up with those of the general population.**

***Results*—The response rate exceeded 75% in a patient population. Changes across the SF 36 questionnaire were associated with self reported changes in health, as measured by the transition question. The questionnaire showed significant improvements in health status for all four clinical conditions, whether in referred or non-referred patients. For patients with suspected peptic ulcer and varicose veins the SF 36 profiles at one year approximate to the general population.**

***Conclusions*—These results provide the first evidence of the responsiveness of the SF 36 questionnaire to changes in perceived health status in a patient population in the United Kingdom.**

(*Quality in Health Care* 1994;3:186–192)

Department of Public Health, University of Aberdeen
Andrew M Garratt, research fellow
Danny A Ruta, lecturer

Health Services Research Unit, University of Aberdeen
Mona I Abdalla, research fellow
Ian T Russell, director

Correspondence to: Mr A M Garratt, Department of Public Health, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB9 2ZD

Accepted for publication 30 September 1994

Introduction

The need for measures of outcome for use in the health service is widely recognised. However, to date most of the focus has been directed towards establishing the validity and reliability of instruments rather than their responsiveness or sensitivity to changes in health over time.[1 2] Since the purpose of an outcome measure is to quantify the effect of health care, any instrument that is intended for use as an outcome measure in the health service must be shown to be responsive to clinically important changes in patients' perceived health.

Although generally agreed methods exist for assessing validity and reliability, there is less consensus in assessing the responsiveness of outcome measures. Previous work has related change scores on outcome measures to external criteria and processes of care. The degree of concordance between change scores and external criteria has been assessed with transition questions,[3 4] which ask the patient or clinician whether there has been a change in health over some given period and in clinical data – for example, erythrocyte sedimentation rate in rheumatoid arthritis.[5] Several studies examined changes in outcome variables after interventions of known efficacy, including total joint arthroplasty[6] and total hip arthroplasty.[7] Different methods have also been proposed for quantifying the responsiveness of outcome measures. These include receiver operator characteristic curves,[4] responsiveness indices,[9] effect sizes,[10] and standardised response means.[6]

The SF 36 health survey questionnaire[11 12] has been shown to be valid and reliable in the United States[12 13] and in populations in the United Kingdom.[14–17] However, only two published studies in the United States examined the ability of the questionnaire to detect changes in health status over time.[7 18] In the first study[7] the 108 item sickness impact profile,[19] previously shown to be responsive by demonstrating significant improvements in patients undergoing joint arthroplasty,[20] was administered with the SF 36 questionnaire preoperatively and postoperatively to patients undergoing total hip arthroplasty. The SF 36 questionnaire was as responsive as the longer sickness impact profile to changes in health status after surgery. In the second study, the SF 36 questionnaire was administered to patients before and after heart valve replacement surgery.[18] Patients' scores were compared with scores for the general population after correcting for sex and age. Before surgery, compared with the general population, the patients had lower scores across the eight scales of the SF 36 questionnaire and scored particularly low on the scales of physical functioning, role limitations, and energy and fatigue. One month after surgery patients'

scores remained below those of the general population on all but the scale of general health perception, and six months after surgery they approximated to those of the general population on all but the scales of role limitations.

If the SF 36 questionnaire is to be adopted as a measure of outcome for use in the NHS its ability to detect changes in health status for patient populations in the United Kingdom needs to be assessed. We report the results of a study in which the questionnaire was administered to a large sample of patients in the United Kingdom with four common conditions – low back pain, menorrhagia, suspected peptic ulcer, and varicose veins – who were followed up for a year. For the eight scales of the questionnaire change scores were calculated and compared with patient responses on a transition question. A standardised measure was used to quantify responsiveness, and patient scores at baseline and follow up were compared with those of the general population.

Methods

SAMPLING AND DATA COLLECTION

The methods of identifying and recruiting patients were as described in the previous paper (p 180)[17] in relation to study 1. In this paper we assessed the responsiveness of the SF 36 questionnaire by sending a follow up questionnaire to patients taking part in study 1 at one year. A random sample of 900 members of the general population, selected from the electoral register for Aberdeen, served as a comparison group; they were sent a similar questionnaire.

STATISTICAL ANALYSIS

Several methods have been proposed for quantifying the responsiveness of outcome measures.[4 6 8 9] The standardised response mean – that is, the mean change in score for a scale, divided by the standard deviation of the change in scores – allows statistically meaningful comparisons to be made between instruments. Higher standardised response means indicate a greater effect or clinically important change, with standardised response means of 0.2, 0.5, and 0.8 or above representing small, moderate, and large clinical changes respectively.[6]

The SF 36 questionnaire contains a transition question which is not used to score any of the eight scales. This question ("Compared to one year ago, how would you rate your health in general now: much better, somewhat better, about the same, somewhat worse, much worse?") was used in this study as a criterion by which to judge the responsiveness of the questionnaire. Such questions are a valid way of measuring changes in perceived health[21] and were used to assess the responsiveness of instruments designed to measure outcomes.[3 4 22 23] For the SF 36 questionnaire to be a valid measure of outcome which reflects perceived changes in health status a significant relation would be expected between changes on the eight scales over the year and the responses to the transition question. Patients indicating an improvement in health on the transition question would be expected to have higher standardised response means across the eight scales than patients who stated that their health remained the same.

Although standardised response means represent a measure of responsiveness of an instrument, a standard is required that facilitates interpretation of scale scores and changes in those scores. To fulfil this purpose normative data can be collected for general or disease free populations.[15 16 24] We compared patient scores on the SF 36 questionnaire at baseline and follow up with data for the general population of Aberdeen after correcting for age, sex, and socioeconomic status. Ordinary least squares regression[25] was used to estimate the effect on each of the eight scales for each of the four conditions, age, sex, and socioeconomic status (housing tenure and age on leaving full time education).[16 24]

Referred patients have a lower perceived health status than those being managed solely in general practice.[16] In this study we compared referred and non-referred patients for changes in their perceived health status over one year. Regression analysis was used to estimate the effect of referral on health status at baseline and follow up for each of the four conditions after correcting for age, sex and socioeconomic status.

The results are presented as absolute and standard scores at baseline for a member of the general population with average characteristics and at baseline and after one year's follow up for members of the four patient groups with the same characteristics. Absolute scores, presented as mean deviations from the scores of the general population, allow comparisons between the patient groups and the general population for each individual SF 36 scale. Standard scores are calculated by dividing the differences between the scores for each patient group with a specific condition and for the general population by the standard deviation of the score for general population. Presented as line graphs, these standard scores allow comparisons between the patient groups and the general population at baseline and after one year's follow up across the entire SF 36 health questionnaire profile.

Results

RESPONSE RATE

The results of the recruitment of patients are described in the previous paper (p 180) in relation to study 1. Of the 1148 patients remaining in the study at one year, 240 failed to respond to the follow up questionnaire, 101 refused to take part, and 32 questionnaires could not be delivered by the post office. Of the 775(67·5%) patients who returned a completed questionnaire at one year, 759 attempted the SF 36 questionnaire at both one year and baseline. Of the comparison sample of 900 members of the general population, 542(60·2%) returned a questionnaire. Their mean age has 47·9 years (range 18–91 years) and 292(53·9%) were female.

*Table 1 Mean changes in SF 36 scales during one year by self reported health transition over one year**

*Reported transition**	*No of subjects*	*Physical functioning*	*Social functioning*	*Role limitation – physical*	*Role limitation – emotional*	*Mental health*	*Energy and fatigue*	*Pain*	*General health perception*
Much better	148	10·50	23·50	44·33	26·57	13·81	21·59	34·11	10·89
Standard deviation		22·91	26·02	44·83	48·70	19·99	18·73	27·50	17·90
95% Confidence interval		6·76 to 14·23	19·23 to 27·77	36·94 to 51·71	18·52 to 34·62	10·51 to 17·10	17·74 to 25·43	29·59 to 38·63	7·99 to 13·80
SRM		0·46	0·90	0·99	0·55	0·69	0·93	1·24	0·61
Somewhat better	173	7·60	12·62	26·84	17·47	7·29	9·31	16·95	3·66
Standard deviation		18·77	24·10	42·35	42·75	16·85	15·51	26·84	14·73
95% Confidence interval		4·78 to 10·42	8·99 to 16·25	20·38 to 33·28	10·92 to 24·02	4·74 to 9·83	6·48 to 12·14	12·88 to 21·01	1·46 to 5·88
SRM		0·40	0·52	0·63	0·41	0·43	0·50	0·63	0·25
About the same	332	2·69	4·82	10·82	6·17	1·39	4·37	10·14	1·33
Standard deviation		16·83	20·80	41·71	43·16	15·72	17·51	23·51	13·47
95% Confidence interval		0·87 to 4·51	2·57 to 7·08	6·24 to 15·39	1·46 to 10·89	−0·32 to 3.10	2·46 to 6·27	7·57 to 12·71	0·13 to 2·78
SRM		0·16	0·23	0·26	0·14	0·09	0·25	0·43	0·10
Somewhat worse	74	−8·89	−9·60	−10·76	−5·71	−2·01	−6·53	0·70	−9·84
Standard deviation		19·77	23·49	36·30	48·14	19·56	19·81	22·18	19·52
95% Confidence interval		−13·47 to −4·31	−15·04 to −4·16	−19·29 to −2·23	−17·19 to 5·77	−6·64 to 2·62	−11·15 to −1·91	−4·57 to 5·87	−14·36 to −5·32
SRM		−0·45	−0·41	−0·30	−0·12	−0·10	−0·33	0·03	0·50
Much worse	22	−16·83	−21·52	−3·57	−7·94	−3·46	−7·14	1·05	−6·35
Standard deviation		23·56	34·40	24·09	39·31	17·90	15·54	23·28	18·46
95% Confidence interval		−27·28 to −6·39	−37·18 to −5·86	−14·54 to 7·39	−25·83 to 9·96	−11·39 to 4·48	−14·22 to −0·07	−9·54 to 11.66	−14·53 to 1·83
SRM		−0·71	−0·63	−0·15	−0·20	−0·19	−0·46	0·05	−0·34

*Question 2 of the SF 36 questionnaire: "Compared to one year ago how would you rate your health in general now?"
SRM = Standardised response mean.
Changes in all eight scales are linearly related to transition scores (F test for linear trend[25]; $p < 0·0001$).

CHANGES IN HEALTH STATUS

Table 1 shows the mean changes on each SF 36 scale by self reported health transition at one year. Of the 749 patients who attempted the transition question, 321(42·9%) perceived their health as better, 332(44·3%) stated that their health had remained the same and 96(12·8%) stated that their health had got worse over the year compared to baseline.

The relation between changes in each of the eight SF 36 scales and the responses to the transition question was highly significant (table 1). For example, patients who perceived their health as much better compared with one year previously showed an average improvement of 10·50 on the physical functioning scale whereas those who perceived their health as much worse showed an average deterioration of −16·83 on this scale. For patients who perceived their health as much better standardised response means representing large changes in health were seen for four of the eight scales and moderate standardised response means for the remainder, with physical functioning on the borderline. For patients who perceived their health as somewhat better standardised response means of a moderate level were seen for four of the scales, the remainder representing small changes. For patients who perceived their health as about the same standardised response means representing a small change were seen for four scales. For patients who perceived their health as being somewhat worse standardised response means representing a small negative change were seen for five of the scales. Finally, for patients who perceived their health as much worse at one year, standardised response means representing a moderate negative change were seen for two of the scales, the remainder representing a small negative change or less (table 1).

To facilitate interpretation of baseline and follow up scores they were compared with data for the general population of Aberdeen. The mean baseline scores for the general population and the mean score deviations for the condition specific groups after correcting

Table 2 Mean SF 36 questionnaire scores for general population and mean score deviations for condition specific groups (mean changes at one year's follow up) after correcting for score age, sex, and socioeconomic status

Group	*No of subjects*	*Physical functioning*	*Social functioning*	*Role limitation – physical*	*Role limitation – emotional*	*Mental health*	*Energy and fatigue*	*Pain*	*General health perception*
General population	542	79·2	78·6	76·5	75·0	73·7	61·2	76·9	68·7
Patients with:									
Low back pain	322	−27·4*(5·4*)	−23·3*(9·5*)	−57·7*(23·8*)	−31·1*(10·6*)	−13·7*(4·4*)	−20·2*(5·6*)	−43·8*(19·0*)	−13·2*(1·0)
Menorrhagia	168	−1·6(−0·0)	−12·3*(8·0*)	−31·3*(14·6*)	−27·9*(15·8*)	−12·3*(6·3*)	−18·5*(11·8*)	−20·9*(9·9*)	−10·9*(3·4¶)
Suspected peptic ulcer	122	+2·7(1·7)	−6·3¶(10·7*)	−15·1*(18·5*)	−6·8(10·7‡)	−7·5*(5·5†)	−11·3*(10·1*)	−23·0*(19·5*)	−7·6*(3·9¶)
Varicose veins	157	−4·1¶(5.0*)	+1·3(2·9)	−14·3*(11·3†)	−8·6*(9·0¶)	−0·0(2·7¶)	−2·9(3·9†)	−7·8*(9·0†)	−1·3(3·1‡)
Referred patients with:									
Low back pain	147	−34·5*(5.9†)	−28·0*(8·7*)	−60·7*(18·1*)	−39·7*(14·4*)	−18·3*(6·3†)	−25·6*(6·4†)	−46·2*(15·0*)	−17·5*(0.5)
Menorrhagia	125	−1·2(−0·0)	−13·8*(8·4*)	−35·4*(16·9*)	−31·2*(16·3*)	−13·7*(7·2*)	−20·9*(13·0*)	−24·3*(12·0*)	−13·0*(5·2†)
Suspected peptic ulcer	58	+1·5(1·4)	−10·3†(13·2*)	−21·5*(17·3‡)	−9·2(12·9¶)	−10·7*(8·2†)	−18·1*(15·4*)	−27·9*(24·8*)	−10·2*(9·0†)
Varicose veins	116	−5·0¶(5·7*)	+0·8(3·5)	−15·5*(13·3*)	−9·3¶(10·8‡)	−0·7(3·5¶)	−4·5¶(5·0†)	−8·8*(10·1*)	−1·4(3·5‡)
Non-referred patients with:									
Low back pain	175	−21·4*(4·9‡)	−19·2*(10·3*)	−55·6*(28·4*)	−23·2*(7·5¶)	−9·8‡(2·8¶)	−15·7‡(5·0†)	−41·9‡(22·2*)	−9·6‡(1·4)
Menorrhagia	43	−3·7(0·0)	−9·6¶(6·6)	−21·5*(7·5)	−15·7¶(14·5¶)	−7·7‡(3·9)	−11·9‡(8·3¶)	−11·5‡(3·6)	−6·8¶(−2·3)
Suspected peptic ulcer	64	+3·8(2·0)	−2·4(8·3*)	−9·7¶(19·5†)	−4·3(8·7)	−4·6(3·0)	−5·2¶(5·2¶)	−18·4‡(14·4*)	−4·9(−0·6)
Varicose veins	41	−1·2(2·5)	+2·5(0·8)	−10·7(4·0)	−4·5(2·2)	2·8(−0·1)	−2·5(−0·2)	−4·4(4·9)	−1·6(1·7)

Mean score for a given patient group can be calculated by adding the mean deviation to the mean score for the general population. Adding the change score in parentheses gives the mean score at 12 months' follow up.
Significant score differences from the general population and significant change scores: ¶$p < 0·05$; ‡$p < 0·01$; †$p < 0·005$; *$p < 0·001$.
Underlined change scores refer to comparisons of change scores for referred and non-referred patients with the same condition (single underlining denotes $p < 0·05$ and bold underlining $p < 0·01$).

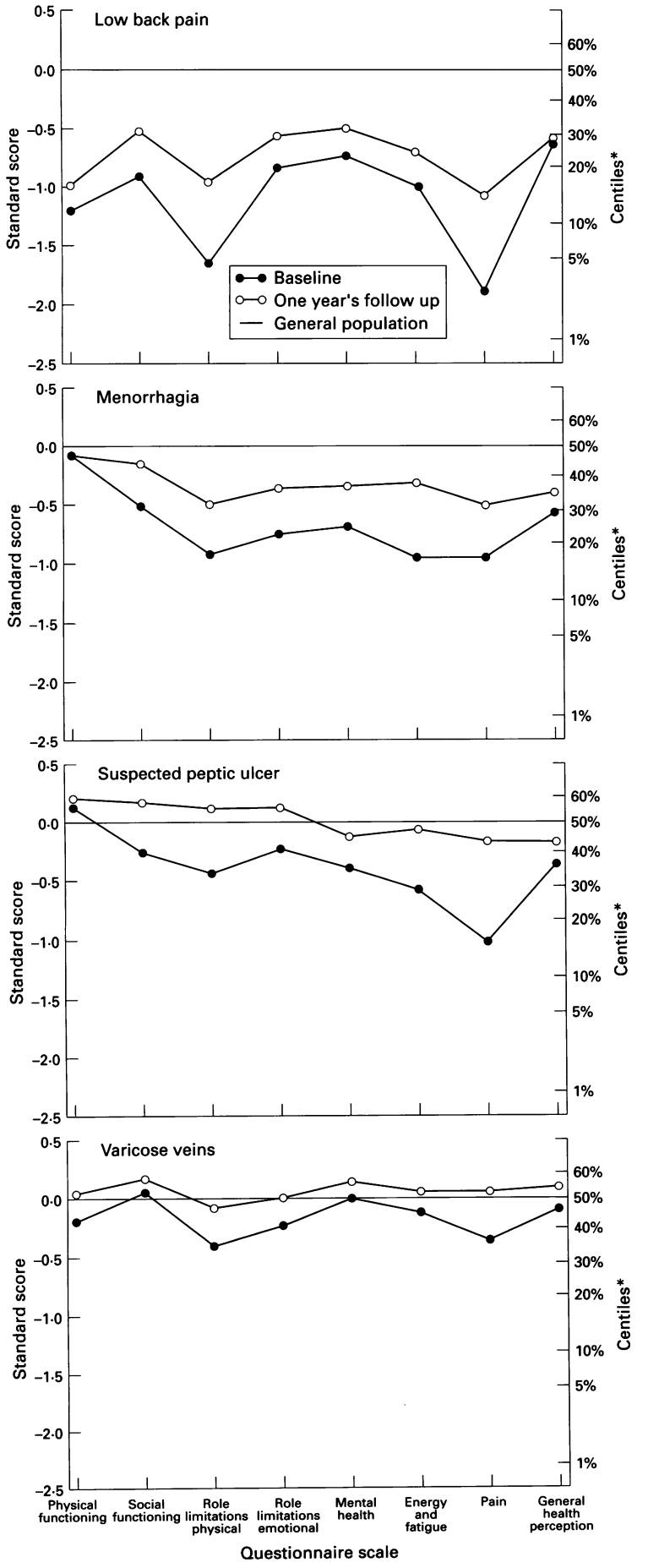

*Fig 1 Mean SF 36 health profiles at baseline and one year's follow up for general population and patients with one of four study conditions (*centiles of general population: for example, X% of general population falls below Xth centile and remaining members fall above that point)*

for age, sex, and socioeconomic status are shown in table 2. The absolute change scores in brackets showed that significant improvements occurred for all four patient groups across seven of the eight SF 36 scales. The scores for patients with back pain improved significantly across seven scales, but they did not show any change on the scale of general health perception. Patients with menorrhagia and peptic ulcer showed significant improvements in scores on all but the physical functioning scale, a scale on which they did not differ significantly from the general population at baseline. Patients with varicose veins showed significant improvements in scores for seven scales but not for social functioning, on which they did not differ significantly from the general population at baseline.

Figure 1 shows the standard scores for an average member of the general population and the same patient with each condition in turn at baseline and one year. The mean for the general population was set at zero on each health scale, allowing comparisons to be made in terms of standard deviations of the general population. For example, patients with low back pain with a mean standard score of about −2 for pain at baseline corresponded to the lowest scoring 2.5% of the general population. The figure serves to highlight the scales on which the patient groups showed the greatest improvement, represented by those areas of health that get closer to the standard score of zero for the general population. The largest areas of improvement included the scales of pain and role limitations attributable to physical problems for low back pain; energy and fatigue and pain for menorrhagia; pain and role limitations attributable to physical problems for suspected peptic ulcer; and pain and role limitations attributable to physical problems for varicose veins. After one year the health profiles for patients with low back pain and menorrhagia still deviated considerably from the health profile of general population whereas the profiles for patients with suspected peptic ulcer and varicose veins more closely approximated to the profile of the general population (fig 1).

Table 2 shows the mean score deviations and change scores for the referred and non-referred patient groups. The same data are presented as standard scores in figure 2. Although significant improvements in SF 36 scores were seen for referred and non-referred patients with back pain and menorrhagia, the health profiles of these two groups of patients at one year was still appreciably lower than the health profile for the general population. However, the health profiles of referred and non-referred patients with suspected peptic ulcer and varicose veins at one year were much closer to the health profile of the general population.

Compared with non-referred patients, patients referred with menorrhagia, suspected peptic ulcer, and varicose veins had a significantly higher level of improvement across the SF 36 scales at one year, for menorrhagia on the scale of general health perception and for suspected peptic ulcer on

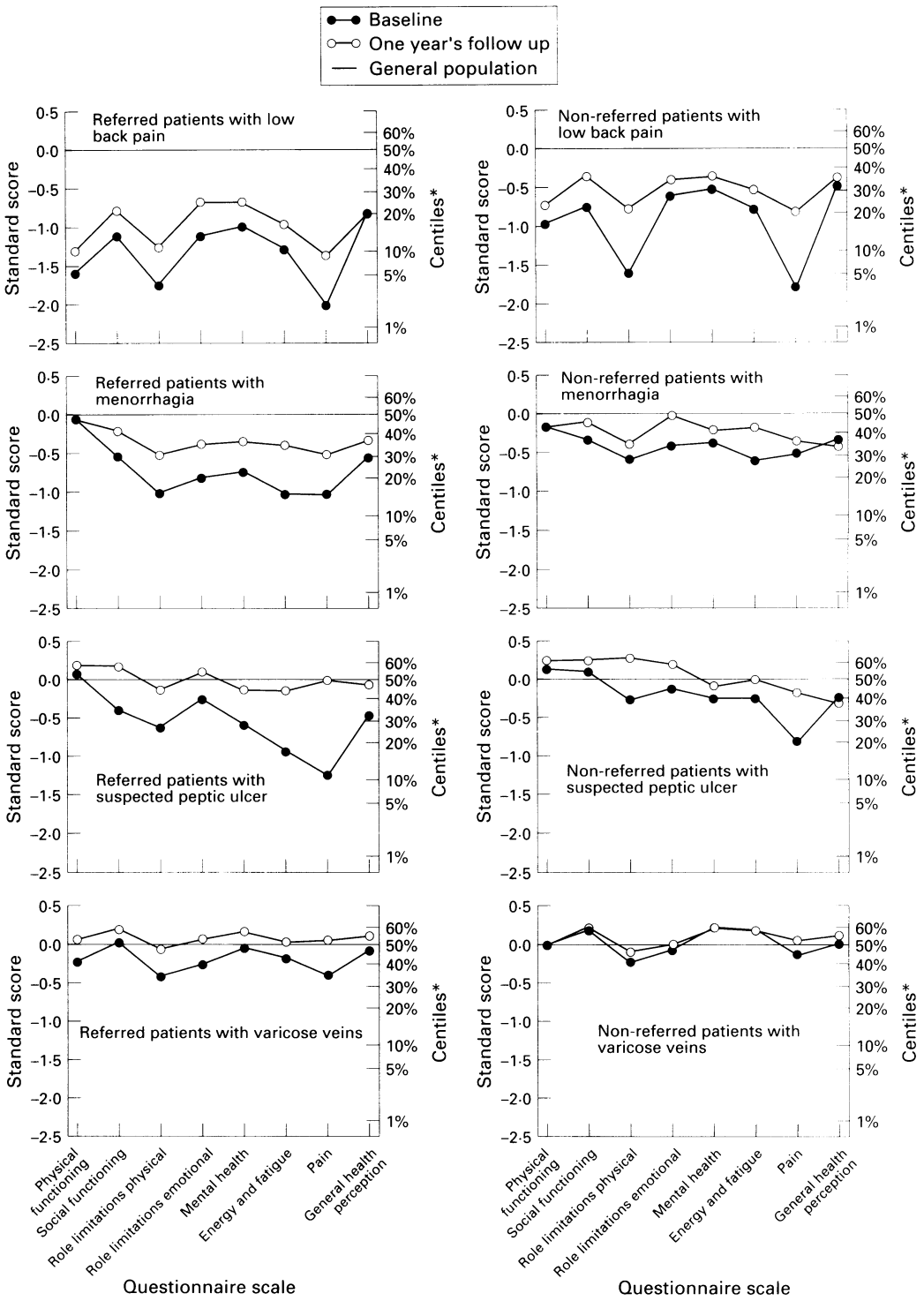

*Fig 2 Mean SF 36 health profiles at baseline and one year's follow up for general population and referred and non-referred patients with one of four study conditions (*centiles of general population: for example, X% of general population falls below Xth centile and remaining members fall above that point)*

the scales of energy and fatigue, pain, and general health perception. For low back pain non-referred patients compared with referred patients improved significantly more according to the scales of role limitations attributable to physical problems and pain.

Discussion

The SF 36 health survey questionnaire has been shown to possess a high level of validity and reliability in populations in the United States.[12 13 26] In the United Kingdom it is a reliable and valid measure of health status in general populations[14 15] and in patient populations.[16] Although some concern has been expressed over the reliability of the SF 36 questionnaire over repeated administrations,[27] we showed in the accompanying article that the questionnaire demonstrates good reliability across all scales when used to monitor health status in groups of patients over time.[17] It is less reliable when used in managing individual patients[17] but is still capable of reflecting clinically important deviations from the norm. If the SF 36 questionnaire is to function as an outcome measure for use in evaluating patient

care and monitoring patients it must also be responsive to changes in health status over time. Results from the United States have proved encouraging, with the SF 36 questionnaire performing as well as the longer sickness impact profile.[7]

In this study changes in all eight SF 36 scales were significantly related to changes in self reported health as measured by the transition question in the questionnaire. Standardised response means,[6] a measure of an instrument's responsiveness to change, were calculated for each SF 36 scale. The group of patients reporting their health as much better on the transition question had the largest standardised response means, followed by those reporting their health to be somewhat better and, lastly, by those reporting their health to be about the same.

To estimate the burden of each particular condition on health status and to assess the extent of any improvement, patients' scores on the eight scales of the questionnaire at baseline and follow up were compared with the scores for the local general population. Across all eight questionnaire scales mean improvements in health status were seen for all four conditions, which were in line with the responses to the transition question. For two conditions – namely, varicose veins and suspected peptic ulcer – the SF 36 health profiles at follow up were close to the health profile for the general population. This finding suggests that for the average patient with suspected peptic ulcer or varicose veins the health outcome at one year was to achieve a standard of health that closely approximates to that experienced by the average member of the general population. For patients with low back pain and menorrhagia changes in scores suggest the outcome is relatively less favourable. Although significant improvements in questionnaire scores were seen for these patients, particularly for pain in patients with low back pain and energy and fatigue in women with menorrhagia, the average patients' score still deviated considerably from that for the average member of the general population, on all eight health scales in the case of low back pain and on all but the physical and social functioning scales in the case of menorrhagia.

Despite evidence of variations in referral patterns[28] referred patients have a lower perceived health status than those being managed within general practice.[16] We showed that improvements in perceived health were greater for referred patients than non-referred patients for three of the conditions under study, the exception being low back pain. The relative lack of improvement in the questionnaire scores for referred compared with non-referred patients with low back pain may have several possible explanations. Referred patients may represent a selected group of patients with back pain who are more likely to have chronic or severe underlying disease and a consequently poorer prognosis or who were referred because their back pain was resistant to treatment. In addition, few treatments for low back pain have been shown to be of clear value.[29]

We emphasise that it has not been possible in a study of this kind to attribute with certainly the health outcomes detected by the SF 36 questionnaire to the medical care received. We were concerned with evaluating the responsiveness of the SF 36 questionnaire and not the effectiveness of treatments. To carry out such an evaluation would require us to relate more precisely the processes of care for individual patients to the outcome achieved, controlling for other variables that may influence health. The alternative would be to conduct a randomised controlled trial of treatment using the SF 36 questionnaire as part of a package of outcome measures. Such a package would also include more traditional clinical indicators and a condition specific measure.[30] For certain conditions, condition specific measures have been shown to be more responsive than the SF 36 questionnaire to small, but clinically significant, changes over time, with greater power to discriminate between patients with very severe or very mild disease.[31] They also provide additional information on symptomatology which may be relevant to clinicians.

For a measure of health status and health outcome to be suitable for routine use within the NHS in a wide variety of clinical settings it must provide information that is valid, reliable, responsive to change, and quick and easy to collect. Our findings here and in the accompanying article suggest that the SF 36 questionnaire is responsive to clinical change and sufficiently reliable[17] for monitoring groups of patients and, for at least four of its scales, individual patients.

In conclusion, the SF 36 questionnaire is responsive to changes in health status over time, even when used to assess health improvements in patients with relatively minor clinical conditions. Taken together with previously published data on validity and reliability, our results provide further evidence for the potential of the SF 36 questionnaire as a routine tool in monitoring and assessment of health outcome in the NHS and as an evaluative instrument in clinical research for a wide range of conditions.

We thank the staff at Inverurie, Portlethen, Rubislaw Place, and Westhill practices for recruiting patients; Jeremy Grimshaw, Jenny Duncan, and Alison De Ville for help with data collection; Elizabeth Russell for helpful comments; and John Ware and his colleagues at the Health Institute of the New England Medical Center for permission to use the SF 36 health survey questionnaire. This research and the Health Services Research Unit are both funded by the Chief Scientist Office of the Scottish Office Home and Health Department; however, the opinions expressed are those of the authors, not of the SOHDD.

1 Tuley MR, Mulrow CD, McMahan CA. Estimating and testing an index of responsiveness and the relationship of the index to power. *J Clin Epidemiol* 1991;**44**:417–21.
2 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993;**20**:417–21.
3 Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res* 1984;**19**:275–89.
4 Mackenzie CR, Charlson ME, DiGioia D, Kelley K. Can the sickness impact profile measure change? An example of scale assessment. *Journal of Chronic Diseases* 1986;**39**:429–38.
5 Jenkinson C, Ziebland S, Fitzpatrick R, Mowat A, Mowat A. Sensitivity to change of weighted and unweighted versions of two health status measures. *International Journal of Health Sciences* 1991;**2**:189–94.

6 Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;**28**:632–42.
7 Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer form health status instruments. *Med Care* 1992; **30**:917–25.
8 MacKenzie CR, Charlson ME, DiGioia D, Kelley K. Can the sickness impact profile measure change? An example of scale assessment. *Journal of Chronic Diseases* 1986; **39**:429–38.
9 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987;**40**:171–8.
10 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**(suppl):S178–89.
11 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
12 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF 36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;**31**:247–63.
13 McHorney CA, Ware JE, Rogers W, Raczek AE, Lu JF. The validity and relative precision of MOS short- and long form health status scales and Dartmouth COOP charts. Results from the Medical Outcomes Study. *Med Care* 1992;**30**(suppl):MS253–65.
14 Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, *et al.* Validating the SF 36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;**305**:160–4.
15 Jenkinson C, Coulter A, Wright L. Short form 36 (SF-36) health survey: normative data for adults of working age. *BMJ* 1993;**306**:1437–40.
16 Garratt AM, Ruta DA, Abdulla MI, Buckingham KJ, Russell IT. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993;**306**:1440–4.
17 Ruta DA, Abdalla MI, Garratt AM, Coutts A, Russell IT. SF 36 health survey questionnaire. I. Reliability in two patient based studies. *Quality in Health Care* 1994; **3**:180–5.
18 Phillips RC, Lansky DJ. Outcomes management in heart valve replacement surgery: early experience. *Journal of Heart Valve Disease* 1992;**1**:42–50.
19 Bergner M, Bobbitt RA, Pollard WE, Martin DR, Gilson BS. The sickness impact profile: validation of a health status measure. *Med Care* 1976;**14**:57–67.
20 Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritic research. *Arthritis Rheum* 1985;**28**:542–7.
21 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Transition questions to assess outcomes in rheumatoid arthritis. *Br J Rheumatol* 1993;**32**:807–11.
22 Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy of diagnostic test performance. *Journal of Chronic Diseases* 1986;**39**:897–906.
23 Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients: the floor phenomenon. *Med Care* 1990;**28**:1142–52.
24 Stewart AL, Greenfield S, Hays RD, Wells K, Rogers MH, Berry SD, *et al.* Functional status and well being of patients with chronic conditions: results from the medical outcomes study. *JAMA* 1989;**262**:907–13.
25 Armitage P, Berry G. *Statistical methods in medical research.* 2nd ed. Oxford: Blackwell, 1987.
26 Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 health survey manual and interpretation guide.* Boston: Nimrod Press, 1993.
27 Sheldon T. Reliability of SF 36 remains uncertain. *BMJ* 1993;**307**:125–6.
28 Roland MO, Coulter AM. *Hospital referrals.* Oxford: Oxford University Press, 1992.
29 Quebec Task Force. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on spinal disorders. *Spine* 1987;**12**(suppl):1–59.
30 Garratt AM, Macdonald LM, Ruta DA, Russell IT, Buckingham JK, Krukowski ZH. Towards the measurement of outcome for patients with varicose veins. *Quality in Health Care* 1993;**2**:5–10.
31 Ruta DA, Garratt AM, Wardlaw D, Russell IT. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994;**19**:1887–96.